

# Development of a computer-based method for a full-length cDNA project and discovery of cis sense-antisense mRNAs with full-length cDNAs

著者	大里 直樹
内容記述	"January 2007"--Cover Thesis (Ph. D.)--University of Tsukuba, (B), no. 2289, 2007.3.23 Includes bibliographical references (p. 120-146)
発行年	2007
URL	<a href="http://hdl.handle.net/2241/91603">http://hdl.handle.net/2241/91603</a>

氏 名 (本籍)	おお さと なお き 大 里 直 樹 (千 葉 県)		
学 位 の 種 類	博 士 (学 術)		
学 位 記 番 号	博 乙 第 2289 号		
学位授与年月日	平成 19 年 3 月 23 日		
学位授与の要件	学位規則第 4 条第 2 項該当		
審 査 研 究 科	生命環境科学研究科		
学位論文題目	<b>Development of a Computer-Based Method for a Full-Length cDNA Project and Discovery of <i>Cis</i> Sense-Antisense mRNAs with Full-Length cDNAs</b> (完全長 cDNA プロジェクトのためのコンピュータによる手法の開発と、完全長 cDNA 配列を用いた <i>Cis</i> sense-antisense mRNA の発見)		
主 査	筑波大学教授	農学博士	深 水 昭 吉
副 査	筑波大学教授	博士 (薬学)	柳 澤 純
副 査	筑波大学助教授	博士 (薬学)	木 村 圭 志
副 査	筑波大学助教授	博士 (農学)	谷 本 啓 司

## 論 文 の 内 容 の 要 旨

ゲノムプロジェクトにおいて、多くの生物種の全ゲノム配列が決定され、ゲノム配列にコードされた遺伝子 DNA の配列や機能の予測が行われてきた。しかし、ゲノム配列からの遺伝子 DNA の配列や機能の予測は、その手法によって結果が大きく異なる。そこで、遺伝子配列や機能をより正確に知るために、完全長の cDNA 配列を決定する、完全長 cDNA プロジェクトが主に日本において始められた。限られた予算、時間、労力で、より多くの種類の完全長 cDNA クローンを収集するためには、重複した完全長 cDNA クローンを cDNA ライブラリーから除く必要がある。重複した完全長 cDNA クローンを除く実験方法が開発されたが、完全に重複を除くには至っていない。そこで本研究では、重複した完全長 cDNA クローンを除くために、完全長 cDNA クローンの端読み配列の情報をもとに、重複した完全長 cDNA クローンを予測し分類するコンピュータの手法を開発した。方法として、完全長 cDNA クローンの配列を決定する前に、完全長 cDNA クローンの 5' end 及び (または) 3' end の配列を決定し、それらの配列の相同性検索を行い、重複した完全長 cDNA クローンを予測した。配列の相同性の度合いに応じた 2 つの基準を考案し、重複した完全長 cDNA クローンを分類した。分類された完全長 cDNA クローンから、完全長 cDNA 配列を決定する代表クローンを選ぶために、各分類に含まれる完全長 cDNA クローンのうち、その全長が最も長いと思われる配列を予測し選択した。マウス完全長 cDNA プロジェクトにおいてすでに配列が決定された完全長 cDNA クローンをを用いて、その 3' end の配列を本方法により新たに分類した結果、従来の方法に比べて、約 30% 多く重複した完全長 cDNA クローンを予測し除くことができることがわかった。本方法をイネ完全長 cDNA プロジェクトに適用し、約 3 年間で 32,127 のイネ完全長 cDNA クローンの配列を決定した。

本研究の 2 番目の目的として、マウス及びイネの完全長 cDNA プロジェクトで得られた完全長 cDNA 配列を用いて、*Cis* sense-antisense mRNA の探索を行った。*Cis* sense-antisense mRNA は、ゲノム上の同じ遺伝子座において、両方のゲノム鎖に位置し、一部相補的な配列を含む。その機能として、遺伝子の発現に

影響し、神経、目、歯、筋肉の形態形成に影響する例が知られている。さらに、X染色体不活化、ゲノムインプリンティング、DNAメチル化、RNAエディティング、オルタナティブスプライシングに影響する例が報告されている。方法として、60,770のマウス完全長cDNA配列と公共データベースにあるマウスcDNA配列をマウスのゲノム配列と比較してそのゲノム上の位置を同定し、*Cis* sense-antisense mRNAの位置関係にあるcDNA配列を検索した。その結果、従来実験により知られていた*Cis* sense-antisense mRNAの数は40程度であったが、本方法により2,481の*Cis* sense-antisense mRNA配列を発見した。またX染色体上では*Cis* sense-antisense mRNAの数が少ないことが分かった。さらに32,127のイネ完全長cDNA配列と公共データベースにあるイネcDNA配列をイネゲノム配列と比較した結果、イネにおいても687の*Cis* sense-antisense mRNA配列を発見した。このことから、*Cis* sense-antisense mRNAが様々な生物種で、今までに報告されていた数よりも非常に多く存在することが示唆された。

近年、タンパク質配列をコードしないNon-protein-coding RNAが遺伝子発現に影響する例が知られてきている。ヒト遺伝子の約20-30%がNon-protein-coding RNAの一種であるmicroRNAのターゲットになることが予測され、遺伝子発現が制御されている可能性があることが報告されている。今回、マウス完全長cDNA配列を解析した結果、11,265のNon-protein-coding RNAが見つかった。しかし、イネ完全長cDNA配列においては、その数はわずか25であった。この傾向は*Cis* sense-antisense mRNAにも当てはまり、マウスの*Cis* sense-antisense mRNAはNon-protein-coding RNAを多く含んでいた。マウス及びイネの完全長cDNA配列の解析から、*Cis* sense-antisense mRNAやNon-protein-coding RNAが多数存在することが明らかになり、RNAレベルの遺伝子発現制御が、従来考えられていたよりも、多く行われていることが示唆された。

## 審 査 の 結 果 の 要 旨

完全長cDNAクローンの収集及びその配列決定は、新規遺伝子の発見や遺伝子機能の解析に重要である。しかし完全長cDNAクローンの網羅的な収集には、多くの手間と費用と時間を必要とする。本論文では、完全長cDNAクローンの収集を効率よく行うために、完全長cDNAクローンの端読み配列の情報をもとに配列が重複した完全長cDNAクローンを予測し除く手法を開発し、本手法をイネ完全長cDNAプロジェクトに用いた。その結果、短期間に多くの種類の完全長cDNAクローンを収集し、その配列を決定することに成功した。さらにマウス及びイネ完全長cDNAプロジェクトから得られた完全長cDNA配列をゲノム配列と比較することにより、機能性RNAの一種である*cis* sense-antisense mRNAsが、従来知られていた数に比べて、きわめて多く存在することを示唆した。またマウス完全長cDNAにコードされている遺伝子の機能を、配列相同性検索などから予測する機能アノテーションを行い、タンパク質配列をコードしていないmRNAであるNon-protein-coding mRNAが多数存在することを明らかにした。

本論文ではモデル生物種の完全長cDNAクローンを網羅的に収集するプロジェクトを世界に先駆けて行うことに貢献し、収集されたクローンの数の多さと決定された配列の精度の高さにおいて、世界的に高い評価を得ており、解析データの信頼性は高い。また、完全長cDNAクローンの配列解析から、従来あまり存在しないと考えられていた、RNAレベルでの遺伝子発現制御が頻繁に行われている可能性を示し、この分野の研究を進展させる先鞭をつけたことに寄与し、高く評価できる。

よって、著者は博士（学術）の学位を受けるに十分な資格を有するものと認める。